Machine Understanding for Interactive Storytelling

Wim De Mulder, Quynh Do Thi Ngoc, Paul van den Broek, and Marie-Francine Moens

KU Leuven, Department of Computer Science
Celestijnenlaan 200A, 3000 Heverlee, Belgium
University of Leiden, Department of Social Sciences
Wassenaarseweg 52, 2333 AK Leiden, The Netherlands

Abstract. This paper describes our research-in-progress which integrates several domains, in particular natural language processing and the development of virtual immersive environments. Our research aims at "bringing a given text to life" via an immersive environment where the user can freely explore the surroundings and increase his understanding of the given text. We describe some important challenges in achieving this goal and outline our current research results. Our work is practically oriented, aiming at fulfilling some societal needs related to education on which we also report.

Keywords: Natural language processing, machine learning, cognitive psychology, digital immersive environments

1 Introduction and outline of the paper

In this paper we present an artificial intelligence challenge that is truly interdisciplinary, covering such diverse domains as natural language processing (NLP), machine learning and the development of interactive computer simulated environments, computer graphics and cognitive psychology. The challenge is to develop a methodology that bridges the gap between natural language and the formal representations of interactive storytelling. Although the domains of natural language processing and the development of immersive digital environments have both seen rapid progress in recent years, their integration is an untouched research domain. Thus the intriguing question: how to create immersive environments that are driven by texts that are written in natural language, but that are not especially written for this purpose? That is, whereas virtual worlds are developed by first handcrafting a knowledge base, the ultimate challenge we present is to consistently transform *arbitrary* texts, not only in terms of content but also in terms of genre (scientific articles, blogs, etc.), to virtual worlds in such a way that these worlds become the main source of information.
The challenge we present here regards a project called MUSE (an acronym for Machine Understanding of interactive StorytElling), which recently started at our university, in collaboration with several universities worldwide, representing

specializations in cognitive psychology, machine learning, and the development of interactive computer simulation environments. The project website is `http://muse-project.eu/`. The presented results were also realized during the EU project TERENCE (An Adaptive Learning System for Reasoning about Stories with Poor Comprehenders and their Educators) (http://www.terenceproject.eu/). In section 2 we shortly describe our project goal and outline some of the main challenges in achieving this goal. Section 3 describes our methodological framework, while in section 4 current research results are outlined. Our research is driven by some societal needs, on which we report in section 5.

2 Translating natural language texts into virtual worlds

2.1 Description of our goal

Our goal is to introduce a new way of exploring and understanding information by "bringing text to life" through 3D interactive storytelling. Stories for children and patient guidelines will be taken initially as input. The choice of stories for children is motivated by the relative simplicity of such stories, while the choice of patient guidelines is motivated by our collaboration with the Haute Autorité de Santé of France[1], an institute that has a large data base with medical guidelines, and the collaboration with the School of Computing of Teesside University, having a lot of expertise with virtual medical environments. The fact that stories for children and medical guidelines have a completely different structure has the advantage that it forces us to develop generally applicable methods.
The given input texts will then be translated into formal knowledge that represents the actions, actors, plots and surrounding world. In a next step this formal knowledge is rendered as virtual 3D worlds in which the user can explore the text through interaction and guided game play.

2.2 Outline of challenges

Incorporating background knowledge One of the main challenges in developing a system that can automatically translate any written text into a digital immersive environment is the incorporation of background knowledge. Although interactive storytelling systems have already been developed, they are limited to a small number of domains in which the developers have already constructed a knowledge base containing a formal representation of the objects, characters, events and processes with which the user can interact. What is missing is the instantaneous incorporation of background knowledge, such that the resulting 3D environment is more than a simple visualization of the text events. As stated in [3], a critical issue in incorporating background knowledge is to circumvent the problem that understanding and incorporating background knowledge requires one to already have a sufficient amount of background knowledge, as even the most detailed encyclopedia articles assume a large amount of common-sense

[1] `www.has-sante.fr/`

knowledge. The challenge is not only to incorporate *correct* background knowledge. As the background knowledge possessed by human beings is very diverse, a crucial question is how to ensure that the 3D interactive environment allows the user to gain the knowledge that is exactly *relevant* to him.

Challenges in natural language processing of texts NLP techniques are necessary to bring natural language closer to 3D immersive environments. We performed experiments on samples of stories for children and patient guidelines, to reveal the main NLP challenges that are encountered during our project. Our main conclusion is that a considerable amount of work needs to be done to bring natural language processing methods to the level that they can be used to fully understand texts in an automatic way. In the context of the proposed goal, especially relevant challenges are:
• Extracting temporal relations. Rendering actions in a narrative text makes the recognition of their temporal relations (e.g., "before") a necessity, a task where improvement of the state-of-the-art is required.
• Extracting spatial relations. Recognizing how the mentioned entities relate to each other spatially in a 3D world is a difficult task that only starts to emerge.
• Recognition of rhetorical relations and rhetorical structure. Examples of such structures are causal and conditional structures, e.g. "if surgery is not suitable in your case, the multidisciplinary team will explain the reasons why and offer you another treatment".
• Coreference resolution, i.e. detecting when multiple expressions in a sentence refer to the same thing. We applied the Stanford Deterministic Coreference Resolution System [12] on our sample texts, resulting in a F-measure [4] of only 67% for the stories for children, and even lower scores for the patient guidelines.

Our methodological framework is shown in Fig. 1. Text is considered as input, which is semantically processed and generic semantic roles (such as actions, actors, temporal and locational expressions) are recognized in the sentences, as well as generic discourse relations such as coreference resolution, and temporal and spatial relations. The semantic labels that are assigned to the text follow the guidelines of annotations generally accepted in the computational linguistic community. In a next step the text, together with its semantic labels is translated into the knowledge representations and planning languages used for virtual reality creation. This step is seen as a machine translation task, augmented with advanced, supervised and unsupervised machine learning of this mapping. As a result the texts can be brought to life and rendered in a 3D-environment.

3 Methodological framework

4 Current research results

4.1 Text annotation tool

The first kind of results we have obtained are in the domain of NLP, more specifically the annotation of texts in a generic way so as to ease the development of a

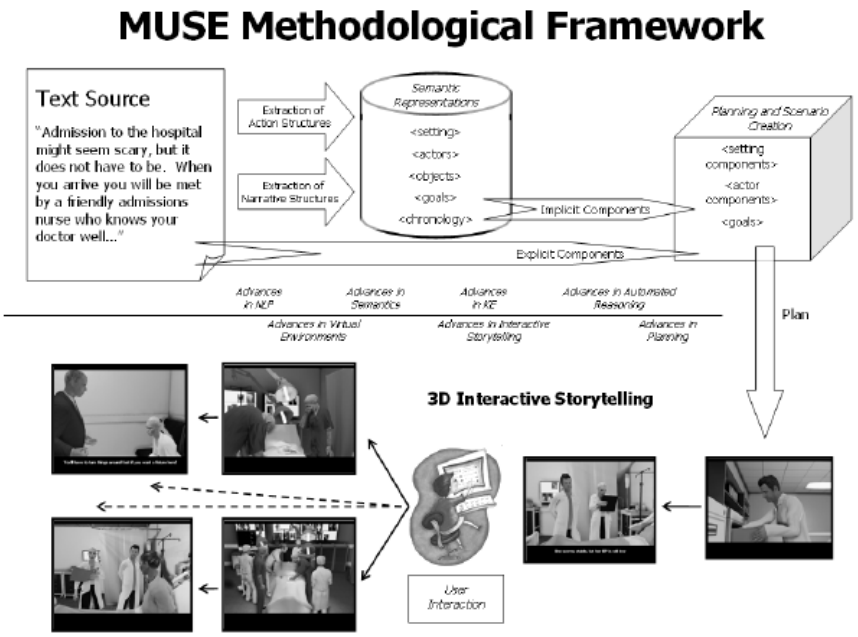

Fig. 1. Methodological framework

mapping strategy from text to virtual environment. We have made advances in the recognition of temporal information [5], the recognition of temporal relations [1], the recognition of spatial relations and spatial relation attributes [9].
This work resulted in a tool[2] for the recognition of events, including their actors and coreferents, and temporal relations in a text. This tool is already successfully evaluated on English stories [1, 6]. The two natural language processors (one in English and one in Italian) analyze flat stories and annotate the factual events and some links between them, e.g., temporal links. A snapshot of some part of a text and the corresponding annotation by our tool is shown in Fig. 2. The figure shows that the annotation tool is able to detect the temporal ordering of the events (e.g. the event where two travellers are on a road is detected as happening before the appearance of a bear) and able to detect semantic roles (e.g. in the event corresponding to the expression 'he observed them', 'he' is recognized as the agent and 'them' as the patient of the event). These first results led to important insights with regard to semantic processing of text.

Temporal ordering is essential for text comprehension and received great interest in the research community (Fig. 2 and 3). In language, temporal ordering is implemented by aspectual and tensed cues. However, in many cases this information is not enough to provide an accurate temporal analysis since most

[2] `http://ariadne.cs.kuleuven.be/TERENCEStoryService/`

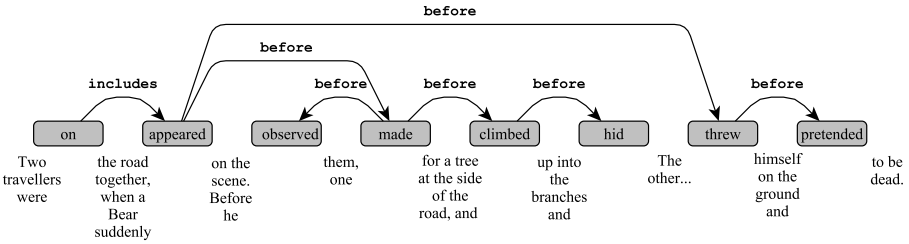

Fig. 2. A timeline for a narrative text. Nodes are events and edges are temporal relations signaled by linguistic cues in the text.

of the cues reside in commonsense knowledge. *Event durations* are very important information for visualization, however, there is very little evidence in text about them. This information humans receive from other sources (e.g. personal experience and observations). The lack of duration information prevents us from exact visual replication of the story plot (and also using TimeGraph [11] as the computational means for timelines), however, for the chosen text genre and text complexity, the visualized story reached a good approximation level of the plot semantics.

Spatial information is a type of semantic annotations vital for visualizations. All visualizations techniques are designed around motions. At the same time spatial information is always present in narratives, but an automated spatial analysis of text remains a far reaching goal. Two important research initiatives addressed spatial annotations in text: STML [13] and MotionML [7]. In this work we used MotionML and the corresponding annotated corpus for identifying motion actions in text. STML, on the other hand, is designed for annotating fine-grained spatial semantics, and seems to be more powerful for consequent qualitative spatial reasoning, but, at this moment, no annotations in STML are available. Spatial annotations are very important for visualizing the initial spatial scene, but similarly to temporal ordering, this information is usually not directly available in text, where personal commonsense knowledge primes this visualization.

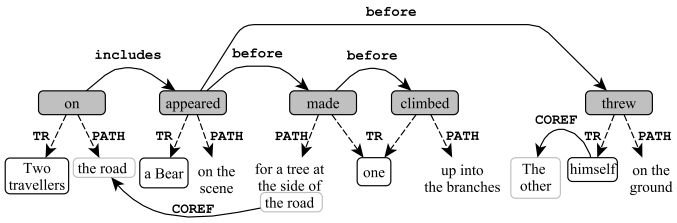

Fig. 3. A story timeline for a narrative text with events considered for actions. Every action represents a motion in the story with argument roles such as *trajector* (TR) and *path* (PATH). Coreferential links are labeled with (COREF).

4.2 Demonstration

The results of the other main line of our research, the creation of immersive environments that are driven by texts, are best evidenced by a demo built in the frame of the TERENCE project [8]. Alice[3] is a 3D graphics prototyping environment developed for teaching and extending computational thinking for students of different age groups and backgrounds. It provides a number of animated and non-animated graphical primitives (persons, animals, characters, vehicles, scenery objects, etc.), which can be placed in a virtual world with customized properties such as color, position opacity, and size. Let us summarize a list of procedural methods applicable to an object in Alice: (i) conversational procedures (`say` and `think`); (ii) orientations (`turn`, `roll`, `turnToFace`, etc.); (iii) positions (`move`, `moveToward`, `moveAwayFrom`, etc.); (iv) appearance (`setOpacity` and `setPaint`). Yet, Alice provides a number of object functional methods, i.e. the methods which return the object's property value (similar to Java getters), for example: `getWidth()`, `getPaint()`, but also spatial properties, such as `isFacing(obj)`, `getDistanceTo(obj)`.

Since animated objects are internally represented as one of the skeletal joint systems, orientation procedures can also be applied to the parts of the skeletal joint system to model. Yet, Alice provides a set of programming statements such as `do_in_order`, `do_together`, `while`, `for_each` and allows one to create visualization/interaction scripts for visualizing complex events and interactions composed by atomic Alice instructions with respect to the user-defined logic.

After actions and actors in the text have been recognized by the NLP pipeline, we use these annotations to populate the virtual world in Alice. A number of assumptions, however, are made: (i) the provided annotations are disambiguated in terms or their meaning, i.e., they are mapped to a single synset in the lexical resource WordNet [10]; (ii) we manually determine visual appearance of characters in Alice; (iii) the spatial setting of the scene is predetermined. The initial spatial layout of the scene is presented in Fig. 4. After the spatial scene has been initialized, we generate Alice procedures from the annotations of actors and actions. The procedure is based on the mapping actions (and their actors) to a set of motion and conversational procedures. The lexical diversity of actions in text is treated by defining the root concepts in WordNet (e.g., `give voice, formulate`#3, and `state, say`#1 for *saying*).

5 Societal benefits

Enhancement of children's understanding of written text has several important advantages, some direct, others more indirect. For example, a major concern on the European agenda is the integration of minority and/or second language groups. An important obstacle for integration and successful participation in a society is the difficulty of communicating. Providing a tool that turns text into action could lead to the development of a broad set of implementations that

[3] `http://www.alice.org`

Fig. 4. The initial story setting with three actors (a Cat and a Hare as Travelers, and a Panda as the Bear). The palm tree represents *a tree at a side of the road.*

would improve communication and, thereby, integration. Our research will also benefit education. International comparisons of educational achievement such as PISA[4] and PIRLS[5] have noted the importance of improving reading abilities. Children in many European countries fall in the middle (or even lower third) of world rankings. Our research will provide a powerful platform by which children can be taught essential strategies for comprehension and learning which goes beyond communication of information [2].
Considering our application to medical guidelines, we notice that patient information and education are major challenges for public health. The main medium for patient education consists of documents produced by health agencies. These documents are often challenging for many patients to understand. Our research will improve the understanding of medical information by patients, thereby not only increasing their knowledge about upcoming surgeries or other treatments, but also reducing patients' anxiety for clinical treatments, as they will have already undergone such treatments virtually in an interactive 3D environment.

6 Conclusion

We presented our research-in-progress which aims at "bringing a given text to life" via the automatic generation of a corresponding virtual environment where the user can freely explore the surroundings and interact with objects and other persons. Current research results are discussed, including a video that demonstrates how a virtual 3D environment can reflect the contents of a story for children. Our research aims at fulfilling some societal needs related to education which are shortly described.

[4] http://www.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html
[5] http://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_IR.html

Acknowledgments

The presented research was supported by the TERENCE (EU FP7-257410) and MUSE (EU FP7-296703) projects.

References

1. Bethard, S., Martin, J.H.: CU-TMP: Temporal relation classification using syntactic and semantic features. In: 4th International Workshop on Semantic Evaluations, pp. 129–132. ACL (2007).
2. Bus, A.G., Neuman S.B.: Multimedia and Literacy Development: Improving Achievement for Young Learners. Taylor and Francis, New York (2009).
3. Gabrilovich, E., Markovitch, S.: Wikipedia-based semantic interpretation for natural language processing. Journal of Artificial Intelligence Research 34, 443–498 (2009).
4. Goutte, C., Gaussier, E.: A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: 27th European Conference on Information Retrieval, pp. 345-359 (2005).
5. Kolomiyets, O., Bethard, S., Moens, M.-F.: Model-portability experiments for textual temporal analysis. In: 49th Annual Meeting of the Association for Computational Linguistics: HLT, pp. 271–276. Association for Computational Linguistics, Stroudsburg, PA, USA (2011).
6. Kolomiyets, O., Bethard, S., Moens, M.-F.: Extracting narrative timelines as temporal dependency structures. In: 50th Annual Meeting of the Association for Computational Linguistics, pp. 87–98. ACL (2012).
7. Kolomiyets, O., Moens, M.F.: MotionML: A shallow approach for annotating motions in text. In: Proceedings of Corpus Linguistics. (2013a).
8. Kolomiyets, O., Moens, M.-F. : Towards animated visualization of actors and actions in a learning environment. In: Proc. 3rd International Workshop on Evidence Based and User-centred Technology Enhanced Learning (2013b).
9. Kordjamshidi, P., van Otterlo, M., Moens, M.-F.: Spatial role labeling: Towards extraction of spatial relations from natural language. ACM Transactions on Speech and Language Processing, 8 (3), article 4 (2011).
10. Miller, G.A.: WordNet: a lexical database for English. Communications of the ACM 38, pp. 39-41 (1995).
11. Miller, S.A., Schubert, L.K.: Time revisited. Computational Intelligence 6(2), pp. 108-118 (1990).
12. Lee, H., Peirsman, Y., Chang, A., Chambers, N.: Stanfords Multi-pass sieve coreference resolution system at the CoNLL-2011 shared task. In: CONLL Shared Task 11 Proc. 15th Conference on Computational Natural Language Learning, pp. 28–34 (2011).
13. Pustejovsky, J., Moszkowicz, J.L.: The qualitative spatial dynamics of motion in language. Spatial Cognition & Computation 11(1), pp. 15-44 (2011).